Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: McWilliams JM, Hatfield LA, Landon BE, Hamed P, Chernew ME. Medicare spending after 3 years of the Medicare Shared Savings Program. N Engl J Med 2018;379:1139-49. DOI: 10.1056/NEJMsa1803388

SUPPLEMENTARY APPENDIX

<u>Table of Contents</u>

Methodological Appendix

Supplemental Tables

Supplemental Figures

A. Beneficiary attribution to ACOs and sensitivity analyses modifying ACO definitions

As previously described,[1] we attributed beneficiaries to ACOs or non-ACO taxpayer identification numbers (TINs) based on which ACO or non-ACO TIN provided the plurality of a beneficiary's office visits with a PCP (defined by specialty codes for internal medicine, family practice, general practice, or geriatric medicine). Office visits used for attribution included the following Current Procedural Terminology codes: 99201-15, 99241-5, G0402, G0438-9, and G0463. We included outpatient specialty consultations (99241-5) in this set because these codes were terminated in 2010, and physicians switched to using 99201-99215. Thus, inclusion of these codes ensured consistency over the study period in the services considered in beneficiary attribution. The code G0463 was created to replace codes 99201-15 for billing for hospital outpatient clinic visits starting in 2014. We included this code, too, to ensure consistency over the study period, per changes in CMS attribution rules.

We considered only office visits with PCPs in the attribution algorithm. Because some ACOs only include PCPs, beneficiaries who only have office visits with specialists (assigned to ACOs in the second step of the CMS ACO attribution algorithm) would be disproportionately attributed to control providers. As we have previously shown, this would introduce imbalance in pre-entry spending levels between ACOs and the control group.[1] We also did not include physician visits in skilled nursing facilities (SNFs) in our attribution algorithm, because their inclusion would similarly introduce imbalance between the ACO and control groups, as most ACOs include no or few providers of SNF care. Physician visits in SNFs were originally included in the CMS attribution rules but were subsequently removed. These modifications we made to the attribution rules introduced some inconsistencies between beneficiaries attributed by CMS to ACOs in the post-entry period and the beneficiaries we attributed to ACOs, but the modifications achieved better balance between comparison groups and thus established a more reasonable counterfactual for difference-in-difference analyses. Of the beneficiaries we attributed to an ACO, 86% were attributed by CMS to an ACO (89% in the 2012 entry cohort, 90% in the 2013 cohort, and 80% in the 2014 cohort). Of beneficiaries attributed to an ACO by both CMS and our methods, the ACO matched in 99.5% of cases.

Of beneficiaries meeting all other inclusion criteria, we excluded 11.1% who had no office visits in a given study year and an additional 12.3% who had office visits only with specialists. A concern with using office visits as the basis for patient attribution is that ACO efforts to increase primary care provision may in turn affect attribution, causing changes in the case mix of patients attributed to ACOs. We found no consistent evidence of differential increases in the number of primary care visits received by ACO patients (Table S4), however, and differential changes in patient characteristics were minimal in almost all cases (Table S1).

In our main approach, we held constant over the study period the set of TINs constituting each ACO contract. We did this to eliminate potential contributions to estimated differential changes in spending from ACO selection of TINs with lower spending when they updated their participant lists in their second or third contract years. Over the study period, however, the practices represented by the TINs included in an ACO's contract could change because of mergers, acquisitions, or changes in the use of specific TINs for billing purposes. In the extreme, an ACO could establish a new TIN to bill under after entering an ACO contract. Although it is not clear conceptually how these changes might be systematically related to MSSP participation and changes in spending, we conducted two sensitivity analyses to gauge the potential bias from these sources of time-varying error.

First, using previously described methods,[1] we added TINs to ACO definitions (lists of participating TINs) to address instances in which TINs were present in claims in some years but not others during the study period. Specifically, we identified TINs in ACO definitions that were missing in at least one of the following years: 2008, 2009, 2013, 2014, or 2015. TINs not present in all of these years either were present at the outset of the pre-entry period but not in the post-entry period, newly appeared in the post-entry period, or disappeared from the beginning to the end of the post-entry period. Because we analyzed a 20% sample of beneficiaries, we expected many small TINs would be intermittently present in the claims because of sampling error. As expected, although 33.4% of ACO TINs were missing in at least one of these years, over 50% of them were solo practitioners, over 75% of them had 2 or fewer physicians billing under them, and over 90% had 5 or fewer physicians. Only 97 of the TINs that were not consistently present in the claims had 20 or more physician NPIs billing under them. We established a procedure for addressing the inconsistencies in ACO definitions introduced by these larger TINs to see if our estimates were affected appreciably. We focused on these larger TINs because their inconsistent presence in the claims was less likely due to sampling error and because they accounted for a substantial proportion of the NPIs billing under TINs that were inconsistently present.

Specifically, for each of the 97 TINs, we identified all NPIs primarily billing under the TIN during the years in which it was present in the claims. We then identified all TINs under which those NPIs primarily billed during the year(s) in which the TIN was absent from the claims. From this step, we determined that 47% of the NPIs billing under the 97 TINs (during the years when the TINs were present) billed under TINs in the alternate period that were also included in the same ACO's definition. Thus, the appearance or disappearance of TINs did not necessarily introduce an inconsistency in an ACO's definition; in many cases, the billing shifted over time among TINs participating in the ACO contract. We identified a subset of the 97 TINs whose NPIs billed primarily under non-ACO TINs in the period when the TINs were absent from the claims. Among those candidate non-ACO TINs, we identified those with stronger connections to the ACO in question based on the number of NPIs billing under them who billed under the ACO's TINs in the alternate period (specifically, we required that 15% of the ACO TIN's NPIs bill under the non-ACO TIN for the non-ACO TIN to be considered part of the same organization). Among non-ACO TINs with weaker connections to the ACO TIN, we identified the practice names using the MD-PPAS file to determine if the non-ACO TIN shared a common name with an ACO TIN. This process led to the addition of 31 TINs to ACO definitions, which added 0.7% more ACO-assigned beneficiary-years to the analysis. As summarized in Table S7, repeating our analysis with these modified definitions of ACOs did not result in appreciably different estimates. In summary, missingness of medium to large-sized TINs in ACO contract definitions over the study period was uncommon. When TINs were added to ACO definitions to address the inconsistencies, results were similar.

Second, to hold constant the physicians included in an ACO contract, we alternatively redefined each ACO contract as the set of NPIs listed for the ACO in the Provider-level RIF and attributed beneficiaries to groups of ACO NPIs instead of groups of ACO TINs, again holding constant the

ACO definition across the study period. Thus, even if a physician was practicing in a different practice before or after the ACO submitted its participant list, the physician would be considered to still be part of the ACO. Likewise if an ACO included a TIN that subsequently acquired another practice, the definition of the ACO would only continue to include the physicians practicing in the ACO practices before the acquisition. Thus, while the procedure described above corrected for missing TINs, holding constant an ACO's set of NPIs corrected for expansions or contractions in the size of an ACO's constituent TINs. This approach also addressed potential systematic selection by ACOs of more efficient physicians over time. While establishing consistency in the physicians constituting an ACO, this approach introduced a different source of error in the alignment of beneficiaries with ACO practices that was greater for years further away from the year in which the ACO contract was defined. As PCPs left ACO practices as part of natural turnover, for example, this approach would continue to consider a patient who followed their PCP as ACO-attributed even though the ACO's primary care practices no longer served the patient. As summarized in Table S7, redefining ACOs as collections of NPIs and holding those definitions constant over the study period did not substantively change our results.

B. Falsification test hypothetically treating non-ACO TINs billing as independent physician groups as ACOs

Using claims from 2015 (approximately the time at which organizational structure of ACOs was determined from ACO and organizational websites), we identified 614 non-ACO TINs with at least 1000 attributed beneficiaries (and thus expected to meet the size criterion for participation in the MSSP of at least 5000 attributed beneficiaries in the full Medicare population). Of these non-ACO TINs, we identified 378 that billed more than 95% of its primary care visits in the independent office setting rather than the hospital-owned setting to approximate the billing practices of ACOs that we categorized as independent physician groups (17 of 203 independent physician group ACOs billed less than 95% of its primary care visits in the independent office setting). The average number of beneficiaries attributed to these non-ACO TINs was1663 beneficiaries per TIN in 2015.

Excluding beneficiaries attributed to ACOs, we then estimated differential changes in spending for patients attributed to these 378 non-ACO TINs, alternately treating them as hypothetical MSSP entrants in 2012, 2013, or 2014. As summarized in Table S5, the differential changes in total Medicare spending were small and not statistically significant, suggesting that we would not expect spending to differentially change systematically (e.g., because of lesser hospital-physician consolidation over the study period) for providers determined to be independent physician groups in the post-entry period. That is, these results suggest that our estimates of differential reductions in total spending from our main analysis were not overestimated.

In these falsification tests, however, there was evidence of differential reductions in spending on outpatient care in the hospital-owned setting and offsetting differential increases in spending on outpatient care in the independent office setting, with no significant net effect on total outpatient spending or total spending (Table S5). This pattern was expected from our post-entry categorization of the large non-ACO TINs as independent physician group, because it held billing in hospital-owned settings at low levels throughout the study period. The magnitude of the differential reductions in these falsification tests suggests that our main analysis may have overestimated the reductions in outpatient spending in hospital-owned settings by 38.5-52.9%, depending on the entry cohort, and that much or all of the differential increases in outpatient spending in the office setting that we estimated in our main analysis would have been observed in the absence of the MSSP. However, our restriction of large non-ACO TINs to those billing over 95% of their primary care visits in the office setting imposed a lower level of billing in the hospital-owned setting than observed among ACOs categorized as independent physician groups (and a higher level of billing in the office setting), as described above. As described below, when we used 2015 entrants for falsification tests, the results suggest overestimation of differential reductions in outpatient spending in hospital-owned settings of approximately 33-44% and no overestimation of the differential increase in outpatient spending in the office setting. Moreover, it is possible that MSSP participation caused less consolidation to occur among participating independent physician groups, availing them of an opportunity to compete with larger health systems by delivering more efficient care.

Thus, taken together, our findings suggest that MSSP participation by independent physician groups has been associated with reductions in spending on outpatient care in hospital-owned settings – e.g., consistent with those ACOs steering patients away from hospital outpatient departments that received higher Medicare reimbursements — and these reductions can only be partly explained by differential exposure to provider consolidation and associated changes in billing from the office to hospital-owned setting. But we cannot quantify the extent to which these reductions in outpatient spending in the hospital-owned setting (or increases in the office setting) can be attributed to MSSP participation as opposed to differences in consolidation that would have occurred in the absence of the MSSP.

To produce a distribution of expected effects in the absence of MSSP incentives, we then drew random samples of non-ACO TINs (with replacement) from the set of 378, treated each draw hypothetically as ACOs, and estimated differential changes in spending from the pre-entry period to 2015, and pre-entry differences in spending trend, for each draw, using the remaining control group as the control group (again excluding beneficiaries attributed to ACOs). We set the size of the random sample to be the average number of independent physician groups in each entry cohort (67) and modeled estimates as if the sample entered in 2012, 2013, or 2014. This produced a distribution of expected effects in the absence of MSSP incentives for each entry year. As displayed in Figure S1, the differential spending reductions estimated for independent physician group ACOs in our main analysis were consistently outside of these expected distributions (and markedly outside of these distributions for the 2012 and 2013 entry cohorts), suggesting that it was very unlikely that these results were due to sampling variation. In contrast, our estimates of pre-entry trend differences (summarized in Table S3) consistently fell within the expected distribution (Figure S1). Results were similar when we drew random samples without replacement (i.e., permutations of non-ACO TINs).

Although differences in pre-entry spending trends in our main analysis (Table S3) were not statistically significant and were small relative to mean total spending (spending trends differed between independent physician group ACOs and the control group by no more than 0.3% of total spending per year [Table 2]), these small trend differences could have explained a nontrivial portion of the estimated effect if they would have continued into the post-entry period in the absence of MSSP participation. In addition, although the pre-entry trend differences from our main analysis fell within the expected distribution in the absence of MSSP participation, they were more negative than most estimates produced by our falsification tests (Figure S1). Therefore, to assess whether we would expect such trend differences that are not considered significant by standard statistical tests to continue in the post-entry period, we estimated the correlation between the pre-entry trend differences and the differential changes across the draws of non-ACO TINs depicted in Figure S1. The correlations were very weak — specifically, 0.01 for 2012 entry tests, -0.01 for 2013 entry, and 0.15 for 2014 entry. These correlations suggest that pre-entry trend differences in the range produced by our falsification tests (which encompassed estimates of pre-entry trend differences like those from our main analysis) would not be expected to predict post-entry trend differences between independent physician group ACOs and the control group.

We conducted another version of this analysis in which we included independent physician group ACOs in the pool of organizations from which the random samples were drawn. Specifically, for a given entry year, the cohort of physician group ACOs entering that year was combined with the same set of large non-ACO TINs. We then drew random samples (with replacement) of organizations from the combined pool (with each organization being either an ACO or non-ACO TIN). The results are displayed in Figure S2. Correlations between pre-entry trend differences and differential changes also were weak in this analysis — specifically, 0.10 for 2012 entry tests, 0.08 for 2013 entry, and 0.17 for 2014.

C. Falsification test hypothetically treating 2015 entry cohort of ACOs as 2012 or 2013 entrants to estimate differential changes in spending expected in 2014 in the absence of MSSP incentives

In our main analysis, we included the 2015 entry cohort as part of the control group, assuming that the first-year effects in 2015 would be minimal. When treating the 2015 cohort instead as a separate entry cohort of ACOs, we estimated the first-year differential change in total annual per-beneficiary Medicare spending (relative to the rest of the control group) to be \$25 (P=0.60) overall for all ACOs in the cohort, supporting this assumption. When stratified by ACO type, we found suggestive evidence of the emergence of a differential reduction for independent physician group ACOs in the 2015 cohort (-\$139; P=0.10). Although not statistically significant at the P<0.05 level, this finding led us to focus on differential changes in spending from the pre-entry period to 2014 in falsification tests treating 2015 entrants as 2012 or 2013 entrants (allowing at least 2 post-entry years).

Table S6 summarizes the results from these falsification tests. As in falsification tests treating large non-ACO TINs as hypothetical ACOs, these analyses treating 2015 entrants as hypothetically earlier entrants also suggested that we likely overestimated spending reductions in outpatient spending in hospital-owned settings (by -\$32/beneficiary per estimates in Table S6) but that we did not meaningfully overestimate *total* spending reductions associated with MSSP participation. See our discussion above of these results in the context of our main results and results of other falsification tests.

These results lend additional robustness to our main results because organizational structure for the 2015 cohort of ACOs was assessed using the same approach as in our main analysis (from websites) rather than relying on billing patterns in claims data, as we did in our other falsification tests. We used descriptions from organizational websites for our main analysis, because many physician practices bill as independent physician groups when part of a larger health system because they are not owned directly by a hospital or hospital system (i.e., a separate system entity can own both hospitals and physician groups). Relying on place of service codes in claims alone would have led to substantial misclassification of organizational structure. Thus, our falsification test treating 2015 entrants as earlier entrants additionally gauged the expected effects of greater provider consolidation in the control group on spending that might not be mediated through changes in place of service coding from office to hospital-owned (e.g., utilization effects of a health system acquiring a physician group practice without a price-increasing change in coding from the office to hospital-owned setting).

D. Resampling analysis to assess importance of ACO organizational structure (independent physician groups vs. hospital-integrated) as a predictor of savings

In addition, we simulated a distribution of effects that would be expected for subgroups of ACOs that share any characteristic or combination of characteristics (or related confounders) that was as prevalent as the organizational characteristic we examined (independent physician group vs. hospital-integrated). Specifically, for each entry cohort, we drew a random sample of ACOs from the cohort (with replacement) that was equal in number to the number of independent physician group ACOs in that cohort. We then estimated differential changes in total Medicare spending per beneficiary for that sample of ACOs, relative to the control group. As displayed in Figure S3, estimates for the subgroups of ACOs that were independent physician groups were consistently in the left tails of these distributions, suggesting that the organizational characteristic we focused on in our analysis was a relatively important predictor of ACO savings when compared with all other possible predictors or sets of predictors that were similarly prevalent among organizations participating in the MSSP. Specifically, differential spending reductions estimated for independent physician group ACOs were greater than spending reductions estimated for 99% of random samples of ACOs drawn from each entry cohort (Figure S3).

For each random draw of ACOs, we also estimated the pre-entry differences in trend between the ACOs and the control group and the correlation between these pre-entry trend differences and the differential changes across draws. As in our analyses of non-ACO TINs described above, these correlations were generally weak and inconsistent across entry years: 0.46 for random draws from the 2012 entry cohort, -0.05 for draws from the 2013 cohort, and 0.11 for draws from the 2014 cohort.

E. Estimation of differences in pre-entry period spending trends and falsification tests treating pre-entry years as post-entry years

To estimate differences in pre-entry trends between each ACO cohort and the control group (reported in Table S3), we added an interaction between the vector "ACO_Cohort" and year, specified as a continuous variable, to the model described in the Statistical Analysis section of the Methods. For each MSSP entry cohort, this term estimated the average change over the pre-entry period in the adjusted difference between ACOs and the control group.

Similar trends (constant differences) in the pre-entry period would support the assumption that differences would remain constant in the post-entry period in the absence of MSSP participation. In the case of outpatient spending for care in hospital-owned facilities, we detected statistically significant differences in pre-entry trends between physician group ACOs and the control group, and as detailed above, falsification tests suggested that we would have expected post-entry differential reductions in this category of spending among physician group ACOs in the absence of MSSP participation (due to a relative increase in such spending in the control group due to hospital-physician consolidation). We detected no significant differences in total spending or in other categories of spending, with some sporadic and inconsistent exceptions (Table S3). Nevertheless, when tests of pre-period trend differences fail to reject the null hypothesis that the trends are the same, we cannot conclude that they are the same. As noted above, a difference in pre-entry trends that is not statistically significant could constitute a meaningful violation of the key assumption of a difference-in-difference analysis (that pre-period differences in outcomes would remain the same in the post-period in the absence of the intervention) if the trend difference continued in the post-period. As summarized in section B, however, we found that differences in pre-entry trends were not predictive of post-entry differential changes when estimated for random draws of physician group ACOs and large non-ACO physician group TINs. This provides compelling evidence that we would not expect differential reductions in total spending for physician group ACOs based on the pre-entry trend differences that we observed.

Consistent with this conclusion, we also found that the serial correlation between consecutive mean year-to-year differential changes in total spending for a given ACO cohort and organizational type was moderately negative (-0.55) over the pre-entry period. In other words, for ACOs of a given organizational type in a given entry cohort, a differential decrease in annual spending relative to the control group from year t to t+1 predicted a subsequent differential *increase* on average from year t+1 to t+2, and vice-versa, over the pre-entry period.

In addition to estimating mean changes in differences between ACOs and the control group over the pre-entry period (trend differences), we also estimated differential changes treating pre-entry years as hypothetical post-entry years (falsification tests). Specifically, for each ACO type and entry cohort we estimated all possible differential changes through the year prior to MSSP entry, alternately using 2009 or 2009-2010 as the hypothetical pre-period. This allowed for estimation of differential changes after 1 or 2 years of hypothetical participation for ACOs in all entry cohorts and estimation of differential changes after 3 years of hypothetical participation for ACOs in the 2013 and 2014 entry cohorts.

The differential changes in total annual Medicare spending estimated in these falsification tests for physician group ACOs are plotted with confidence intervals in Figure S4. The estimates ranged from -$94 per beneficiary to $111 per beneficiary, none differed significantly from zero, and there was no evidence of increasingly negative differential changes as the number of hypothetical years of participation increased from one to three. In contrast, study estimates displayed in Figure 1 (rearranged in Figure S5 to be similarly ordered from most positive to most negative from top to bottom) were consistently negative, all but one estimate were more negative than the most negative estimate produced by the falsification tests, and all but one of the second and third year effects were statistically significant. In addition, the size of the estimates consistently increased with more years of participation (Figure 1 and S5).

F. Accounting for multiple testing in analyses of primary outcome

In a sensitivity analysis, we used the Hochberg procedure to account for multiple testing in analyses of our primary outcome (6 tests for total spending in 2015, 1 for each of the 2 categories of ACOs in each of the 3 entry cohorts). The first step of the Hochberg procedure is to order the P values for k tests from largest to smallest: $P_1 \geq P_2 \ldots \geq P_k$. The corresponding null hypothesis H_{0k} and all null hypotheses H_{0j} for j>k are rejected if $P_k \leq \alpha/k$, where α is the type I error, equal to 0.05. The adjusted P-value for each test is $P_{adjustedk} = \min\{P_1, 2P_2, \ldots kP_k\}$ for k = 1, 2,… K.

When applied to our primary analyses, the ordered P values are: 0.78, 0.14, 0.009, 0.005, <0.001, <0.001. As the procedure moves through these P values, the null hypothesis is rejected for the third through sixth test because 0.009<0.05/3. Thus, the statistical significance of the tests is not altered by the adjustment. The adjusted P values are, in order, 0.78, 0.24, 0.027, 0.02, <0.005, <0.005.

G. Sensitivity Analyses

Results of additional sensitivity analyses are presented in Table S7 with descriptions of the analyses in corresponding notes below the table. Below we discuss our approach to addressing potential differential changes in the composition of patient populations served by ACOs or in the composition of physicians billing under ACO TINs.

Changes in Patient Populations

The extent to which any differential changes in patient characteristics contributed to our findings was examined in analyses estimating differential changes in observable patient characteristics (Table S1) and comparing results with vs. without adjustment for those characteristics (Table S7). For HCC risk score and CCW condition indicators (key time-varying characteristics that could be affected by changes in coding practices), we compared estimates adjusted for risk scores and condition indicators based on recent (prior year) vs. earlier diagnoses (recorded before ACO contracts altered incentives to code more intensively).

To address potential bias from differential compositional changes in the patient populations served by ACOs and the control group, an alternative approach one might consider would be to add beneficiary-level fixed effects to the models we estimated to control for time-invariant characteristics of beneficiaries. We opted against this strategy in our primary analysis because adding beneficiary-level fixed effects would introduce additional sources of potential bias and compromise the interpretability of results. Specifically, adding beneficiary fixed effects would effectively restrict estimation of the differential changes in spending to beneficiaries who are in the sample and eligible for assignment to an ACO or non-ACO provider in both the pre-period and in 2015 (the post-period year of interest), because observations in both periods are required to estimate effects based on within-beneficiary changes. This restriction effectively excludes a high proportion of beneficiaries. For example, of all beneficiaries in our sample, only 38.9%

were in the sample for at least one year from 2009-2011 and in 2015. Because this sample differs systematically from other beneficiaries present in only one of the two periods, one would not be able to determine if any impact of the fixed effects on spending estimates is due to removal of bias vs. effect modification (effects of ACO contracts being different for the included vs. excluded groups).

More importantly, the difference-in-difference estimates would become confounded by any fixed differences between ACOs and the control group in the prevalence of, or practice patterns for, the groups excluded from estimation nonrandomly with respect to time (i.e., the groups appearing in the sample in only the pre- or post-periods but not both). These groups include, for example: beneficiaries who died, permanently enrolled in Medicare Advantage (MA), or transitioned to a long-term care setting during the pre-period; and those who disenrolled from MA into fee-for-service (FFS) Medicare or became newly eligible for Medicare in the post-period. The remaining group present in both periods (the group supporting estimation with beneficiary fixed effects) is thus constrained in ways that systematically differ in the pre- and post-periods. For example, in a beneficiary fixed effects model, beneficiaries are included if they die or transition permanently to a nursing home in the post period (as long as they have an office visit in the year of death or transition) but not if they do so in the pre-period (because they would then not be in the post-period). Similarly, beneficiaries could be included if they disenroll from MA in the pre-period but not the post-period, etc.

Thus, the differential change estimates would reflect any fixed differences between ACOs and control group in spending for beneficiaries in the excluded groups or in the prevalence of the excluded groups. Those differences would no longer be differenced out because the beneficiary fixed effects exclude the counterfactuals for the (collectively large) subgroups that are asymmetrically present in the pre- or post-period, thereby contributing to differential changes.

There are many mechanisms through which this source of bias might play out. For example, ACOs and the control group might differ in their annual mortality rates. With beneficiary fixed effects in the model, that difference would contribute to a post-period difference in spending but not a pre-period difference in spending. As another hypothetical example, recent MA disenrollees, who have been shown to be more costly in the period immediately after disenrolling from MA (e.g., consistent with health shocks requiring greater choice of providers), might be disproportionately in the control group because they keep their PCPs, who might be in mostly MA-oriented physician groups and not ACO-participating groups. Since the beneficiary fixed effects mean that the recent disenrollees from MA can only be in the pre-period in order to contribute to estimation, we would observe a differential increase in spending in the ACO group (a differential reduction in the control group as the disenrollees's spending declines after their health shocks). Without the beneficiary fixed effects, recent MA disenrollees are allowed to contribute to differences in both the pre- and post-periods, so that any related fixed difference in spending between the ACO vs. control group is differenced out.

While one might attempt to address these sources of potential bias by making additional exclusions of decedents, MA enrollees/disenrollees, beneficiaries transitioning to long-term care settings, etc., the remaining sample would be even smaller and even more different from the full sample (with much lower spending). Moreover, the additional exclusions would only partially address the sources of time-varying confounding introduced by beneficiary fixed effects. The inclusion of beneficiary fixed effects would still mean that beneficiaries contributing to estimation would be systematically different (older, sicker, etc.) in the post-period than in the pre-period. Thus, if ACOs treat older and sicker patients more or less intensively, the differential change in per-beneficiary spending would get bigger or smaller as the proportion of older and sicker patients rises (even if the change in proportion is the same for both groups). Including beneficiary fixed effects is therefore analogous to requiring a stronger common shocks assumption. The beneficiary fixed effects ensure that there are shocks to both comparison groups (both get older and sicker with commensurately sharp increases in spending), and if the groups are different (which they are, by definition, because they are treated by different providers), the same shock might affect spending differently in the different groups. Using the full sample without beneficiary fixed effects establishes greater stability in age, disease burden, etc. within each group over time.

The main point is that the assumptions of a difference-in-difference analysis are more likely to hold if groups are similar in the pre- and post-periods. Paradoxically, adding beneficiary fixed effects may reduce comparability. It is also possible that ACOs might affect death or hospice use, for example, raising concerns about bias introduced by conditioning on such events in defining a cohort.

Our approach achieved excellent balance across many observable patient characteristics (minimal differential changes in Table S1). Thus, we are not concerned about differential changes in beneficiary traits when beneficiary fixed effects are not included. Given our concern about the non-random effective sample reductions when beneficiary fixed effects are used and the high degree of sample balance when they are not, we made the pre-specified decision to include all beneficiaries served by ACO and non-ACO providers in the analysis and to consider alternative strategies to address compositional changes in the patient population related to ACO participation only if we observed evidence of such changes based on the many patient characteristics we could measure (which we did not).

In a sensitivity analysis, we nevertheless explored the effects of limiting our sample to a cohort of beneficiaries meeting inclusion criteria for our analysis in both the pre-period (we specified at least 1 from 2009-2011) and the post-period (2015), excluding beneficiaries who died, enrolled in Medicare Advantage, or transitioned to a nursing home or hospice at any point during the study period. Only 35% of beneficiaries in our full sample remained after these exclusions, and mean spending in the pre-period was 37% lower for this cohort than for our full sample. After restricting the sample to beneficiaries in this cohort, we estimated differential reductions in spending associated with MSSP participation for each of the 3 MSSP entry cohorts of ACOs. The estimates were at most 33% smaller in magnitude than those from our main analyses (the largest difference was for the 2012 entry cohort where the differential change in spending was -$203 per beneficiary vs. -$302 in our main analysis). Because mean spending was 37% lower after the exclusions ($6102 vs. $9649 annually), this means that estimates from this cohort analysis were not smaller in relative terms (e.g., -3.3% vs. -3.1% for the 2012 entry cohort). We then added beneficiary fixed effects to models, and the estimates were not meaningfully changed. While these results provide further evidence that our main results were not due to compositional changes in the patient populations served by ACOs, we caution against conclusions based on these additional analyses for the reasons described above.

Changes in ACO Providers

As discussed above in Section A, we held the collections of TINs and CCNs defining each ACO constant over the study period so that changes in the TINs or CCNs participating in ACO contracts would not contribute to estimates of differential changes in spending. Per MSSP rules, ACO benchmarks should be adjusted for any changes in the TINs or CCNs included in ACO contracts, so ACOs should not have strong incentives to select TINs or CCNs with lower spending for inclusion. Because the clinicians (NPIs) billing under a TIN change over time as physicians enter or exit practices or as practices merge or expand, we alternatively defined ACOs as collections of NPIs, holding constant the sets of NPIs over the study period. When holding constant the NPIs comprising each ACO, estimated differential spending reductions were somewhat greater for the 2012 and 2013 entry cohorts of ACOs and somewhat smaller for the 2014 entry cohort of physician group ACOs when compared with our main results (Table S7).

These findings suggest that selection of more efficient physicians into ACO TINs did not explain the spending reductions associated with MSSP participation. In theory, ACOs in the MSSP have an incentive to change the physicians comprising participating TINs in favor of more efficient physicians over time, since their benchmarks would not be updated to reflect such changes. However, in practice, such selection behavior is likely to be challenging for ACOs to implement effectively. First, ACOs would have to know which physicians are more or less efficient, which requires distinguishing variation in practice patterns at a physician level from variation in their patients' characteristics. Because risk adjustment for observable characteristics is invariably inadequate, because PCP practice patterns are likely to account for only a small proportion of patient-level variation in *total* spending, and because physician-level spending fluctuates substantially from random changes in patients and patient needs, reliably making this distinction may not be possible. At the very least, it would require a high level of strategic sophistication and capacity for data analytics that most ACOs, particularly smaller physician group ACOs, may not possess. Moreover, the high degree of stability in patient characteristics we observed (i.e., the consistently minimal differential changes in Table S1) suggests that ACOs would have had to perfectly distinguish physician efficiency from patient risk in selecting physicians for inclusion in ACO TINs; an imperfect distinction would result in observably lower risk patients from selection of ostensibly more efficient physicians with lower adjusted spending. Thus, because we do not observe evidence of differential changes in patient characteristics favoring lower risk patients, an alternative explanation that the spending reductions were partially due to selection of more efficient physicians into ACO TINs seems implausible.

Second, to alter the composition of physicians billing under a TIN, an ACO would have to encourage inefficient physicians to leave their practices, hire new more efficient physicians (with ex ante knowledge of their relative efficiency), or somehow shift billing of inefficient physicians to a non-ACO TIN used by the same organization without changing where physicians practice.

One might also wonder whether ACOs achieved spending reductions by shifting their patients to more efficient member physicians (i.e., without changing the composition of NPIs comprising the ACO). We do not consider such a strategy as a gaming strategy or as invalidating savings achieved by ACOs but do view it (conceptually) as an unlikely explanation for the bulk of the savings because it, too, would require accurate knowledge of the relative efficiency of member physicians. To investigate the potential contribution of such a mechanism, we added NPI fixed effects to our models after re-attributing beneficiaries to specific NPIs (as opposed to groups of NPIs in ACOs) so that all beneficiaries were attributed to a single NPI in each year. In this analysis, we also held constant over the study period the NPIs considered to be part of each ACO. Adding the NPI fixed effects changed our estimates modestly and inconsistently across entry cohorts and ACO type, with changes in estimates ranging approximately from -$75 per beneficiary (greater savings) to +$75 per beneficiary (lesser savings).

We would note, however, that the impact of adding NPI fixed effects is difficult to interpret for reasons analogous to the problems we describe above for models with beneficiary fixed effects. With NPI fixed effects in the model, estimation of differential changes is based on within-NPI changes and thus effectively excludes NPIs that have attributed beneficiaries only in the pre-period or in the post-period. Of all NPIs (PCPs) with at least one attributed beneficiary in 2009-2011 or in 2015, for example, only 57% were present in both periods. The 43% of NPIs excluded from estimation accounted for 17.4% of beneficiaries in the sample and included physicians exiting the workforce or Medicare in the pre-period and new physicians joining the workforce or Medicare in 2015 (they also included NPIs with few Medicare patients who thus intermittently appear in claims for a random 20% of beneficiaries). Thus, if the proportion or practice patterns of exiting or entering physicians differed between ACOs and the control group, difference-in-difference estimates would be confounded by fixed differences that would be differenced out of estimates when NPI fixed effects are omitted from the model.

Changes in care delivery by ACOs also could change the NPI accounting for the most office visits without affecting ACO assignment (i.e., assignment to a collection of TINs or NPIs). Such changes could systematically affect higher- or lower-risk patients, causing NPI fixed effects to introduce bias into estimates of differential spending changes. For example, ACOs have adopted the use of annual wellness visits (AWVs) more than non-ACO providers. AWVs largely substitute for office visits that would otherwise be billed as 99201-15 visits and may be delivered by specific physicians or nurse practitioners (who may be newly hired for that purpose) disproportionately to certain types of patients. Thus, while AWV adoption might not cause much change in ACO assignment, it could cause changes in assignments to specific NPIs (which are necessary in models with NPI fixed effects) that are systematically related to patient or clinician characteristics. Including NPI fixed effects in regression models in addition to restricting ACO definitions to NPIs participating at the outset of contracts could therefore affect estimates in ways that are challenging to interpret.

Effect modification might also contribute to different results when NPI fixed effects are included. For example, ACOs might more easily influence the practice patterns of new physicians who have recently completed their training. Since NPI fixed effects effectively exclude new physicians in the post-period from the estimation, differential changes in spending estimated based on other NPIs would be smaller, whereas an assessment of savings would ideally include such a mechanism.

Finally, we would not expect the practice patterns of PCPs (i.e., the efficiency of the PCP to which beneficiaries are assigned in NPI-specific attribution) would explain much of the changes in spending in acute care, and particularly post-acute care, which were more likely the result of organizational strategies to manage utilization (e.g., shorten SNF length of stay) and to substitute primary care for acute care. To the extent such strategies are implemented for all of an ACO's patients or executed by non-physician staff, one would not expect changes in the physician composition of ACO TINs to account for their effects on spending.

For these reasons, we pre-specified as our primary strategy an approach in which we defined ACOs as collections of TINs or CCNs, allowing changes in the clinician membership of TINs and CCNs as long as we found no evidence of favorable selection at the beneficiary level (i.e., as long as differential changes in patient characteristics were minimal, which they were). Nevertheless, we found no clear evidence of ACOs favoring more efficient clinicians for inclusion in ACO's originally participating TINs as the physician membership of those TINs evolved over the post-entry period, or of ACOs shifting volume within TINs to more efficient clinicians. Specifically, estimates were not substantially and consistently changed by redefining ACOs as groups of NPIs or by additionally including NPI fixed effects in models.

In summary, we have structured our primary analysis to minimize the impact of changing patient or provider traits. Moreover, we believe inclusion of beneficiary or physician fixed effects may paradoxically generate more bias because they effectively systematically exclude a non-random set of patient or providers from the estimation. That said, our findings are robust to inclusion of beneficiary or provider fixed effects, suggesting the theoretical debate is largely inconsequential.

Table S1. Differential changes from the pre-entry period to 2015 in the characteristics of patients served by ACOs, as compared with the control group, by organizational type and entry cohort of ACOs*

Patient characteristic	Unadjusted sample mean in pre-entry period†	Differential change from pre-entry period to 2015 for ACOs vs. control group (N=4,327,280 beneficiaries in 2015)					
		Hospital-integrated ACOs (N=132 ACOs)			Independent physician group ACOs (N=203 ACOs)		
		2012 entry cohort	2013 entry cohort	2014 entry cohort	2012 entry cohort	2013 entry cohort	2014 entry cohort
Age (yr)	72.2 ± 11.9	0.0	0.1	0.1	0.0	0.2	0.1
Female sex (%)	58.5	-0.1	-0.3	0.1	-0.4	-0.2	-0.3
Race or ethnic group§ (%)							
Non-Hispanic white	83.5	0.0	-0.2	0.8	-0.3	-0.1	-0.5
Non-Hispanic black	8.5	0.0	0.2	-0.7	0.4	0.0	0.1
Hispanic	4.7	0.0	-0.1	-0.2	0.1	0.2	0.2
Other	3.2	0.0	0.1	0.1	-0.2	0.0	0.1
Medicaid recipient (%)	15.3	-0.1	-0.1	-0.3	-0.5	0.0	0.3
Disability was original reason for Medicare eligibility (%)	21.9	-0.3	-0.4	-0.3	0.3	-0.2	-0.3
End-stage renal disease (%)	0.9	0.0	0.0	0.0	0.1	0.0	0.0
Nursing home resident in prior year (%)	1.4	-0.1	0.0	0.0	-0.6	0.2	0.0
CCW conditions,¶ no.							
Through prior year	5.71 ± 3.19	-0.02	0.01	-0.02	-0.01	0.06	-0.02
Through 3 years prior**	4.52 ± 3.20	-0.01	0.03	-0.02	-0.01	0.03	-0.02
HCC risk score‖							
Based on claims in prior year**	1.234 ± 1.060	-0.006	0.005	-0.003	0.017	0.033	0.004
Based on claims 3 years prior	1.070 ± 0.887	-0.001	-0.001	-0.004	0.005	0.009	-0.001
ZCTA-level characteristic							
% Below federal policy level	9.2	0.0	0.0	0.0	0.1	0.0	0.1
% With high school diploma	75.4	-0.2	-0.1	0.1	-0.2	0.0	-0.1
% With college degree	19.4	-0.2	-0.1	0.0	-0.3	0.1	0.0

*Plus-minus values are means ± SD. The control group consisted of beneficiaries attributed to non-ACO providers. The 2012 entry cohort included 114 ACOs, the 2013 entry cohort included 106 ACOs, and the 2014 entry cohort included 115 ACOs. Means and percentages were adjusted for geography to reflect comparisons within hospital referral regions. ZCTA denotes ZIP Code tabulation area.

[†]In the analyses the pre-entry period differed for each entry cohort, but for the purpose of describing the study sample in this table, years 2009-2011 were used to calculate a single mean for each characteristic.

[§]Race or ethnic group was determined from Medicare Master Beneficiary Summary Files.

[¶]Chronic conditions from the Chronic Conditions Data Warehouse (CCW) included 27 conditions: acute myocardial infarction, Alzheimer's disease, Alzheimer's disease and related disorders or senile dementia, anemia, asthma, atrial fibrillation, benign prostatic hyperplasia, chronic kidney disease, chronic obstructive pulmonary disease, depression, diabetes, heart failure, hip or pelvic fracture, hyperlipidemia, hypertension, hypothyroidism, ischemic heart disease, osteoporosis, rheumatoid arthritis or osteoarthritis, stroke or transient ischemic attack, breast cancer, colorectal cancer, endometrial cancer, lung cancer, prostate cancer, cataracts, and glaucoma. Analytic models included indicators for all 27 conditions and indicators for the presence of multiple conditions ranging from 2 to 9 or more conditions. Counts of conditions included all conditions except cataracts and glaucoma.

[‖]Hierarchical Condition Categories (HCC) risk scores are derived from demographic and diagnostic data in Medicare enrollment and claims files, with higher scores indicating higher predicted spending in the subsequent year. For each beneficiary in each study year, we assessed the HCC score based on enrollment and claims data in the prior year, two years prior, and three years prior, in each case requiring continuous enrollment in fee-for-service Medicare in the study year and the year of claims used to calculate HCC scores.

**For analyses of CCW condition indicators and HCC scores derived from earlier claims, we limited the sample to a subgroup of beneficiaries who were also continuously enrolled in fee-for-service Medicare 3 years prior to the study year.

Table S2. Pre-entry differences in spending and utilization between ACOs and control group, by MSSP entry cohort*

Spending or utilization measure	Unadjusted sample mean in pre-period	All ACOs			Hospital-integrated ACOs			Independent physician group ACOs		
		2012 entry cohort	2013 entry cohort	2014 entry cohort	2012 entry cohort	2013 entry cohort	2014 entry cohort	2012 entry cohort	2013 entry cohort	2014 entry cohort
Annual per-beneficiary spending ($)										
Total	9,649	139	31	32	98	90	99	189	-56	-25
Total acute inpatient care	3,411	42	31	48	61	79†	63	20	-41	36
Total outpatient care	3,069	-50†	-1	-16	-50	25	24	-48	-43	-53†
Independent office	1,717	37	-48	-27	-21	-144†	-83†	107†	98†	24
Hospital-owned facilities	1,352	-86†	47	11	-29	169†	106	-155†	-141†	-77†
Total post-acute care	1,177	92	10	-7	33	9	-22	163	12	8
Home health care	634	11	11	2	26†	-8	20	-7	39†	-14
Post-acute	109	-3	3	4	-1	0	6	-5	7†	1
Outpatient	525	14	8	-2	27†	-8	14	-2	32	-15
Durable medical equipment	317	-4	5	5	-6	13†	16†	-3	-6	-5
Hospice	172	25	-2	5	12†	2	4	40	-8	5
Annual per-beneficiary utilization, no.										
Hospitalizations	0.350	0.002	0.004	0.002	0.004	0.008†	0.000	-0.001	-0.001	0.003
Hospitalizations for ACSCs	0.044	0.000	0.001	0.000	0.001	0.002†	-0.001	-0.001	-0.001	0.000
Emergency department visits	0.468	-0.022†	-0.012	-0.015†	-0.016†	-0.001	-0.011†	-0.028†	-0.029†	-0.019†
Post-acute facility stays	0.201	0.077†	0.004	0.002	0.008†	0.003	0.005†	0.007	0.007	0.000
Days in post-acute facility	2.29	0.14	0.04	-0.02	0.01	0.05	-0.03	0.28	0.02	0.00
Primary care visits	4.27	0.10	0.47	0.15	-0.17†	-0.07	-0.13	0.43†	1.24	0.41†
Proportion of admissions followed by readmission within 30 days of discharge, %	17.58	-0.09	-0.08	-0.10	-0.17	0.05	-0.17	-0.02	-0.22	0.01

*Adjusted for geography and patient characteristics as in the main analysis
†P<0.05 for test of difference from zero, not adjusted for multiple testing

Table S3. Pre-entry differences in spending and utilization *trends* between ACOs and control group, by MSSP entry cohort*

Spending or utilization measure	Unadjusted sample mean in pre-period	All ACOs			Hospital-integrated ACOs			Independent physician group ACOs		
		2012 entry cohort	2013 entry cohort	2014 entry cohort	2012 entry cohort	2013 entry cohort	2014 entry cohort	2012 entry cohort	2013 entry cohort	2014 entry cohort
Annual per-beneficiary spending ($)										
Total	9,649	-3	-5	8	18	15	34	-33	-32	-14
Total acute inpatient care	3,411	2	-8	8	12	-1	25[†]	-13	-17	-7
Total outpatient care	3,069	-12	-5	-7	-19[†]	2	-1	-3	-15	-13
Independent office	1,717	3	-7	-1	-1	-11[†]	-9	6	-1	6
Hospital-owned facilities	1,352	-14[†]	2	-6	-18[†]	13	8	-9	-13[†]	-19[†]
Total post-acute care	1,177	11	8	-1	16	10	11[†]	3	5	-11[†]
Home health care	634	-3	-5	-1	5	-2	-3	-13	-10	0
Post-acute	109	0	-2[†]	0	2	0	-1	-2	-4[†]	0
Outpatient	525	-3	-3	-1	3	-2	-2	-12	-6	-1
Durable medical equipment	317	-3	-1	-1	-4	1	0	-3	-5	-2
Hospice	172	1	1	3[†]	0	0	2	2	2	4
Annual per-beneficiary utilization, no.										
Hospitalizations	0.520	-0.001	0.000	0.001	0.001	0.000	0.002[†]	-0.004	0.000	0.000
Hospitalizations for ACSCs	0.044	0.000	0.000	0.000	0.001	0.000	0.001[†]	-0.001	0.000	0.000
Emergency department visits	0.468	-0.002	-0.003[†]	-0.003[†]	-0.005	-0.004[†]	-0.002	0.001	-0.001	-0.003[†]
Post-acute facility stays	0.201	-0.001	0.000	0.000	0.000	-0.001	0.001	-0.003	0.001	0.000
Days in post-acute facility	2.29	-0.01	0.02	0.01	0.03	0.03	0.03[†]	-0.06	0.00	-0.02
Primary care visits	4.27	-0.03	-0.01	0.01	-0.01	-0.01	0.01	-0.06	-0.01	0.02
Proportion of admissions followed by readmission within 30 days of discharge, %	17.58	0.06	-0.08	0.03	0.17	-0.20[†]	0.11	-0.05	0.10	0.00

*Adjusted for geography and patient characteristics as in the main analysis
[†]P<0.05 for test of difference from zero, not adjusted for multiple testing

Table S4. Differential changes from the pre-entry period to 2015 in Medicare spending for patients attributed to ACOs, as compared with the control group, by organizational type and entry cohort of ACOs[*]

Spending or utilization measure	**Unadjusted sample mean in pre-entry period**[†]	**Differential change from pre-entry period to 2015 for ACOs vs. control group**											
		Hospital-integrated ACOs						**Independent physician group ACOs**					
		2012 entry cohort		**2013 entry cohort**		**2014 entry cohort**		**2012 entry cohort**		**2013 entry cohort**		**2014 entry cohort**	
		Estimate	P value	Estimate	P value	Estimate	P value	Estimate	P value	Estimate	P value	Estimate	P value
Annual per-beneficiary spending ($)													
Total	9,649	-169[‡]	0.005	-18	0.78	88	0.14	-474[‡]	<.001	-342[‡]	<.0001	-156	0.009
Total acute inpatient care	3,411	-41	0.32	10	0.76	28	0.39	-137	0.008	-127[‡]	0.002	-69	0.03
Total outpatient care	3,069	-41[‡]	0.050	16	0.51	66[‡]	0.01	-47	0.14	-82	0.002	-34	0.15
Independent office	1,717	41	0.02	-18	0.43	6[‡]	0.77	49	0.01	-6	0.75	39	0.050
Hospital-owned facilities	1,353	-82[‡]	<.001	33	0.26	61[‡]	0.09	-95[‡]	<.0001	-76	0.001	-72	<.001
Total post-acute care	1,177	-64	<.001	-16	0.45	23	0.22	-174[‡]	0.02	-56[‡]	0.03	-31	0.14
Home health care	634	-10	0.29	8	0.30	12[‡]	0.049	-28	0.04	-41[‡]	0.003	0	0.98
Post-acute	110	2	0.54	4	0.02	1	0.70	2	0.53	-5	0.08	1	0.69
Outpatient	525	-12	0.15	4	0.58	11[‡]	0.04	-31	0.01	-36[‡]	0.008	-1	0.97
Durable medical equipment	317	5	0.26	-10[‡]	0.07	-12[‡]	0.01	-5	0.56	-6	0.37	-13	0.03
Hospice	172	-3	0.64	-10	0.10	8	0.24	-31	0.14	6	0.51	4	0.62
Annual per-beneficiary utilization, no.													
Hospitalizations	0.350	-0.003	0.36	-0.001	0.88	0.008	0.009	-0.011	0.005	-0.009[‡]	0.005	-0.005	0.13
Hospitalizations for ACSCs[§]	0.044	0.000	0.97	-0.001	0.22	0.001	0.32	0.000	0.90	-0.002	0.10	0.001	0.54
Emergency department visits[¶]	0.468	-0.007	0.20	-0.008	0.16	-0.010	0.02	-0.017[‡]	0.02	-0.024[‡]	<.0001	-0.015[‡]	0.002
Post-acute facility stays	0.201	-0.005	0.005	-0.001	0.72	0.003	0.10	-0.015[‡]	<.0001	-0.012[‡]	<.0001	-0.003	0.25
Days in post-acute facility	2.29	-0.04	0.09	0.00	0.98	0.06	0.04	-0.28[‡]	0.06	-0.11	0.01	-0.04	0.17
Primary care visits[‖]	4.27	-0.01	0.69	-0.13	<.001	0.00	0.99	0.06	0.25	0.05	0.33	0.08	0.06
Proportion of admissions followed by readmission within 30 days of discharge, %	17.58	0.00	0.99	0.44	0.01	0.72[‡]	0.001	0.04	0.86	0.09	0.77	-0.59	0.003

[*]As the primary purpose of the secondary analyses was to decompose the observed effect on total spending into the contributions from its various spending and utilization components, rather than to estimate effects on additional outcomes, P values were not adjusted for multiple testing. See section F of the Supplementary Appendix for adjustment of primary analyses of total spending for multiple testing. PCP denotes primary care physician, and ACSCs ambulatory care-sensitive conditions.

[†]In the analyses, the pre-entry period differed for each entry cohort, but years 2009-2011 were used to calculate a single mean for each measure in the table.

[‡]Test of change in the differential change from the first full post-entry year to 2015 was statistically significant from zero ($P<0.05$) without adjustment for multiple testing. These tests were conducted for differential changes in 2015 that differed significantly from zero. For tests of our primary outcome, total spending, adjustment for multiple testing using the Hochberg procedure did not alter the statistical significance of these tests.

[§] Hospitalizations for ambulatory care–sensitive conditions (ACSCs; conditions for which appropriate ambulatory care could potentially reduce the need for inpatient care) included two measures that are included in ACO contracts (hospitalization for congestive heart failure and chronic obstructive pulmonary disease or asthma) as well as hospitalizations for diabetes or cardiovascular disease, conditions that are the focus of many other quality measures included in the contracts. Specifically, these included hospitalization for uncontrolled diabetes, short-term complications of diabetes, long-term complications of diabetes, amputation of the foot or leg, hypertension, and angina without procedure. Unlike all other outcomes, hospitalizations for ACSCs were affected by the transition of diagnosis codes to the 10^{th} revision of the International Statistical Classification of Diseases and Related Health Problems in October of 2015, because they were defined based on diagnosis codes recorded in inpatient claims. To address this inconsistency across years, we analyzed hospitalizations for ACSCs occurring in the first 9 months of each year.

[¶]Annual counts of emergency department visits include those that were not followed by hospitalization so that emergency department visit and hospitalization counts were mutually exclusive.

[‖]Primary care visits include the office visits with primary care physicians in independent office, hospital-owned, or safety-net settings.

Table S5. Falsification test hypothetically treating large non-ACO TINs as MSSP entrants in 2012, 2013, or 2014

Annual per-beneficiary spending measure, $	Pre-entry difference in trend between non-ACO TINs treated as ACOs and remaining control group	P value	Differential change from pre-entry period to 2015 for non-ACO TINs treated as ACOs vs. remaining control group	P value
Non-ACO TINs treated as 2012 entrants				
Total	18	0.32	-19	0.56
Total outpatient	-16	0.01	2	0.89
Independent office	-1	0.90	53	<.0001
Hospital-owned facilities	-16	<.001	-50	<.001
Non-ACO TINs treated as 2013 entrants				
Total	-8	0.52	-5	0.87
Total outpatient	-18	<.0001	13	0.37
Independent office	2	0.47	50	<.0001
Hospital-owned facilities	-20	<.0001	-38	0.002
Non-ACO TINs treated as 2014 entrants				
Total	-5	0.57	-3	0.91
Total outpatient	-13	<.001	18	0.18
Independent office	5	0.051	45	<.0001
Hospital-owned facilities	-18	<.0001	-28	0.02

Table S6. Falsification test hypothetically treating 2015 entry cohort of ACOs as 2012 or 2013 entrants to estimate differential changes in spending expected in 2014 in the absence of MSSP incentives

Annual per-beneficiary spending measure, $	Pre-entry difference in trend between 2015 entry cohort and control group				Differential change from pre-entry period to 2014 for 2015 entry cohort of ACOs vs. control group			
	All ACOs in 2015 cohort (N=89 ACOs)	**P value**	**Independent physician group ACOs in 2015 cohort** (N=24 ACOs)	**P value**	**All ACOs in 2015 cohort** (N=89 ACOs)	**P value**	**Independent physician group ACOs in 2015 cohort** (N=24 ACOs)	**P value**
Treating 2015 cohort as 2012 entrants								
Total	19	0.49	-18	0.75	93	0.09	-14	0.89
Total outpatient	0	0.98	6	0.75	14	0.89	-25	0.38
Independent office	6	0.43	16	0.26	8	0.64	7	0.68
Hospital-owned facilities	-6	0.36	-10	0.36	6	0.74	-32	0.09
Treating 2015 cohort as 2013 entrants								
Total	39	0.06	63	0.18	65	0.20	-76	0.45
Total outpatient	-4	0.55	8	0.55	18	0.34	-29	0.31
Independent office	0	0.99	11	0.33	10	0.51	3	0.88
Hospital-owned facilities	-4	0.46	-3	0.73	8	0.65	-32	0.08

Table S7. Results of sensitivity analyses

Analysis	Hospital-integrated ACOs			Independent physician group ACOs		
	Difference in pre-entry trend between ACOs and control group	**Differential change in total per-beneficiary spending from pre-entry period to 2015 for ACOs vs. control group**	**P value**	**Difference in pre-entry trend between ACOs and control group**	**Differential change in total per-beneficiary spending from pre-entry period to 2015 for ACOs vs. control group**	**P value**[**]
2012 entry cohort						
Main results	18	-169	0.005	-33	-474	<0.001
Not adjusted for patient characteristics	10	-172	0.01	-27	-495	0.002
TINs added to ACO definitions to address inconsistencies in presence of TINs in claims[*]	10	-161	0.01	-33	-484	<0.001
ACOs redefined as constant sets of NPIs[*]	28	-203	0.002	-52	-489	0.001
ACOs redefined using 2015 definitions[‡]	36	-147	0.02	-36	-480	0.002
Propensity-score weighted to balance characteristics between ACOs and control group within each HRR and year[§]	12	-182	0.002	-14	-455	<0.001
Claims from/through 3 years prior used to derive HCC scores/CCW condition indicators	-27	-202	0.004	-125	-413	0.005
2015 MSSP entry cohort removed from control group[¶]	15	-156	0.01	-31	-462	0.001
Adjusted for ACO's level of 2008 spending relative to the regional average and level of spending in ACO's service area relative to the national average[‖]	-10	-167	0.03	20	-404	<0.001
Including office visits with nurse practitioners and physician assistants in beneficiary attribution	17	-170	0.004	-18	-479	<0.001
2013 entry cohort						
Main results	15	-18	0.78	-32	-342	<0.001
Not adjusted for patient characteristics	18	-31	0.65	-38	-362	<0.001
TINs added to ACO definitions to address inconsistencies in presence of TINs in claims[*]	11	-27	0.67	-29	-342	<0.001
ACOs redefined as constant sets of NPIs[*]	21	-88	0.10	-19	-387	<0.001

ACOs redefined using 2015 definitions[‡]	7	22	0.72	-53	-351	<0.001
Propensity-score weighted to balance characteristics between ACOs and control group within each HRR and year[§]	13	-21	0.72	-45	-352	<0.001
Claims from/through 3 years prior used to derive HCC scores/CCW condition indicators	-21	-16	0.84	-19	-188	0.03
2015 MSSP entry cohort removed from control group[¶]	21	-16	0.80	-27	-320	<0.001
Adjusted for ACO's level of 2008 spending relative to the regional average and level of spending in ACO's service area relative to the national average[ǁ]	5	-28	0.65	-16	-372	<0.001
Including office visits with nurse practitioners and physician assistants in beneficiary attribution	24	64	0.27	-31	-394	<0.001
2014 entry cohort						
Main results	34	88	0.14	-14	-156	0.009
Not adjusted for patient characteristics	21	69	0.23	-23	-182	0.003
TINs added to ACO definitions to address inconsistencies in presence of TINs in claims[*]	34	90	0.13	-14	-153	0.01
ACOs redefined as constant sets of NPIs[*]	10	37	0.49	-37[†]	-99	0.13
ACOs redefined using 2015 definitions[‡]	30	76	0.20	-25	-199	0.001
Propensity-score weighted to balance characteristics between ACOs and control group within each HRR and year[§]	27	92	0.08	-15	-190	0.002
Claims from/through 3 years prior used to derive HCC scores/CCW condition indicators	13	63	0.42	-32	-159	0.01
2015 MSSP entry cohort removed from control group[¶]	35[†]	100	0.10	-12	-159	0.008
Adjusted for ACO's level of 2008 spending relative to the regional average and level of spending in ACO's service area relative to the national average[ǁ]	26[†]	60	0.28	-12	-148	0.02
Including office visits with nurse practitioners and physician assistants in beneficiary attribution	32[†]	219	0.002	-22	-178	0.008

*See description in section A of the Supplementary Appendix above. In addition, when redefining ACOs as sets of NPIs and holding them constant over time, and after altering the attribution algorithm to attribute each beneficiary to a specific NPI (as opposed to a group of NPIs in an ACO), we found that adding NPI fixed effects to models did not substantially affect our estimates (see section G).

†Pre-entry trend difference statistically significant from zero at P<0.05 level, without adjusting for multiple testing

‡We used definitions from 2015, again holding them constant over the study period

§We used a propensity score weighting technique to balance the distribution of beneficiary characteristics between the ACO and control group within each HRR and year.[1,2]

¶To ensure sufficient control group in areas with high MSSP participation, in our main analysis we included all beneficiaries attributed to ACOs entering the MSSP in 2015 as part of the control group, as we expected their first-year effects on spending would be minimal. We conducted sensitivity analyses testing that assumption (Table S6 above) and excluding 2015 entrants from the analysis (corresponding row in Table S7 above).

‖Using previously described methods, we adjusted comparisons of hospital-integrated and independent physician group ACOs for baseline spending levels that also have been found to predict spending reductions in the MSSP.[1,3] Specifically, for each ACO we assessed mean 2008 spending for control group beneficiaries in its service area and categorized ACOs according to whether their local 2008 spending was higher or lower than the average among all ACOs. We also categorized ACOs according to whether 2008 Medicare spending for their attributed beneficiaries was higher or lower than spending for control group beneficiaries in their service area. We assessed baseline spending levels in 2008, before the start of our study period, to minimize contributions from regression to the mean; as previously demonstrated, this approach effectively that goal. For the entire study sample, including both hospital-integrated and physician group ACOs, we then fit a model that included interactions between the ACO_Cohort$_k$×Post$_t$ indicators and: indicators for ACO organizational type, an indicator for having baseline spending above the regional average, and an indicator for being in a higher spending area. As in prior work, we found that ACOs with higher baseline spending for their area and ACOs in higher spending areas had greater spending reductions than ACOs in the corresponding lower spending categories, but these differences were not consistently statistically significant and not as large as in prior work, suggesting that ACOs with lower baseline spending began to achieve spending in later years. Adjustment for these associations between baseline spending and spending reductions did not substantively affect differences in spending reductions between hospital-integrated and physician group ACOs (Table S7 above).

[**]In another sensitivity analysis, we clustered variance estimation at the HRR level for all beneficiaries (vs. clustering at the ACO level for ACO-attributed beneficiaries and at the HRR level for the control group). Clustering at the HRR level for all beneficiaries changed confidence intervals minimally in the 2013 and 2014 cohorts but led to narrow confidence intervals for the 2012 cohort of physician group ACOs. Thus, P values for our main results are more conservative.

Figure S1. Differential changes in total per-beneficiary Medicare spending from pre-entry period to 2015 and pre-entry differences in trend estimated for 100 random draws of large non-ACO TINs that billed as independent physician groups, hypothetically treated as 2012, 2013, or 2014 MSSP entrants

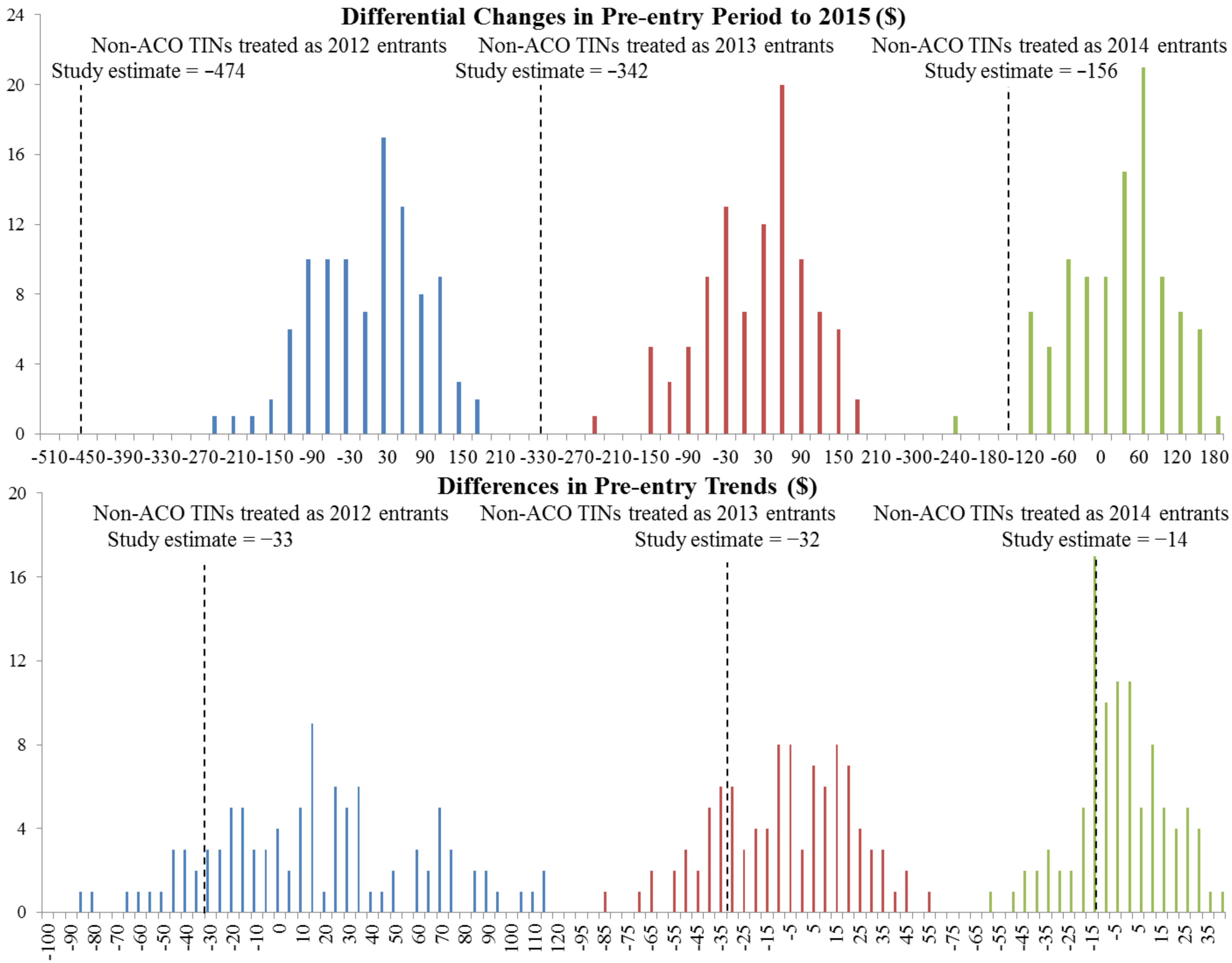


Figure S2. Analysis presented in Figure S1 except with independent physician group ACOs mixed with non-ACO TINs as pool of organizations from which random samples were drawn

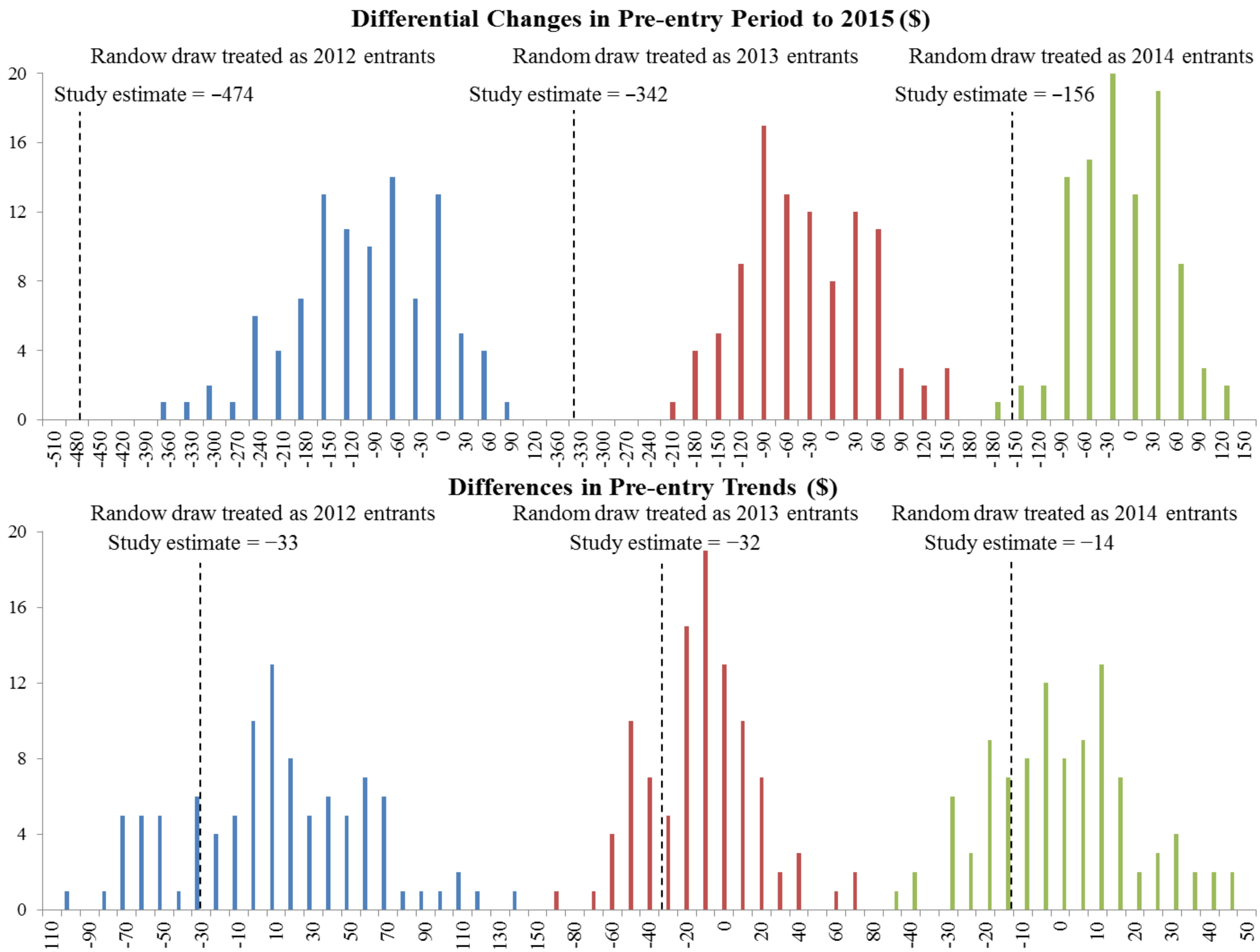


Figure S3. Distribution of estimates for 100 random draws of ACOs from each entry cohort and position of estimates for independent physician group estimates in the distribution

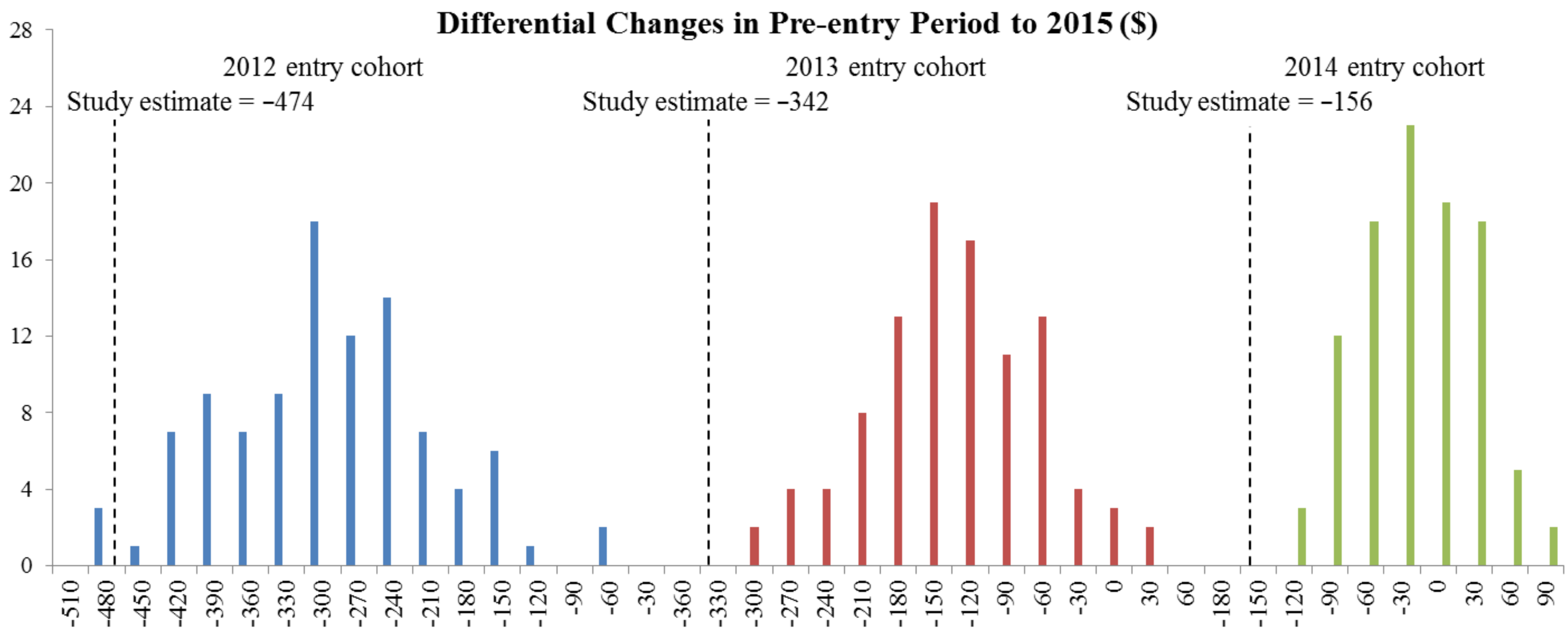
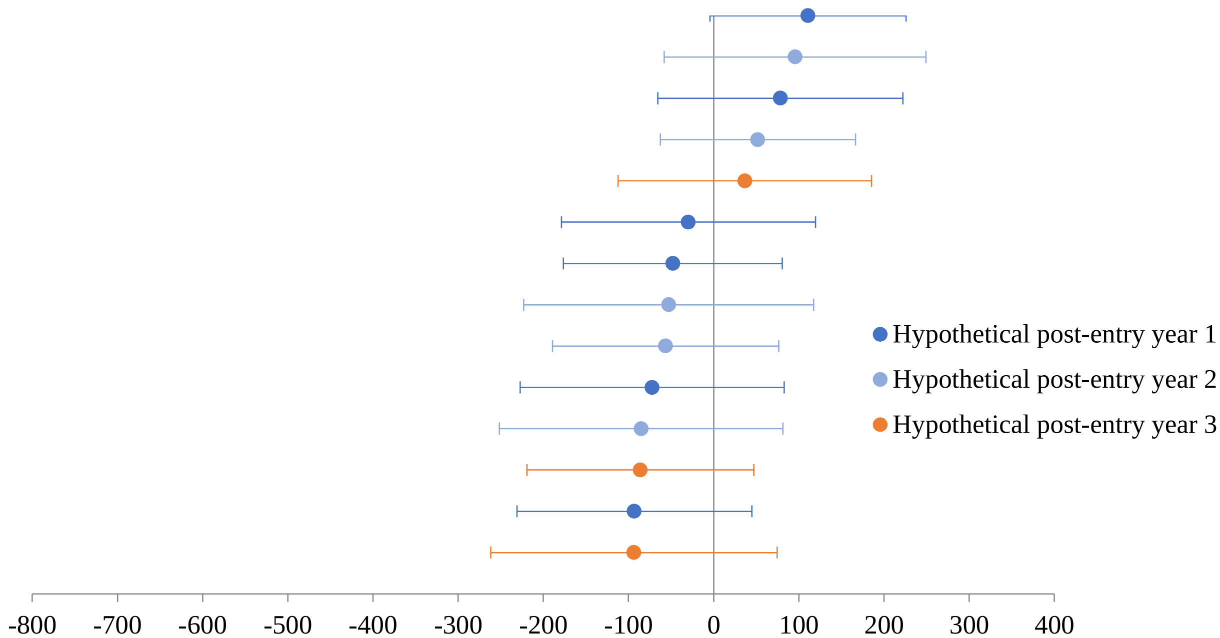


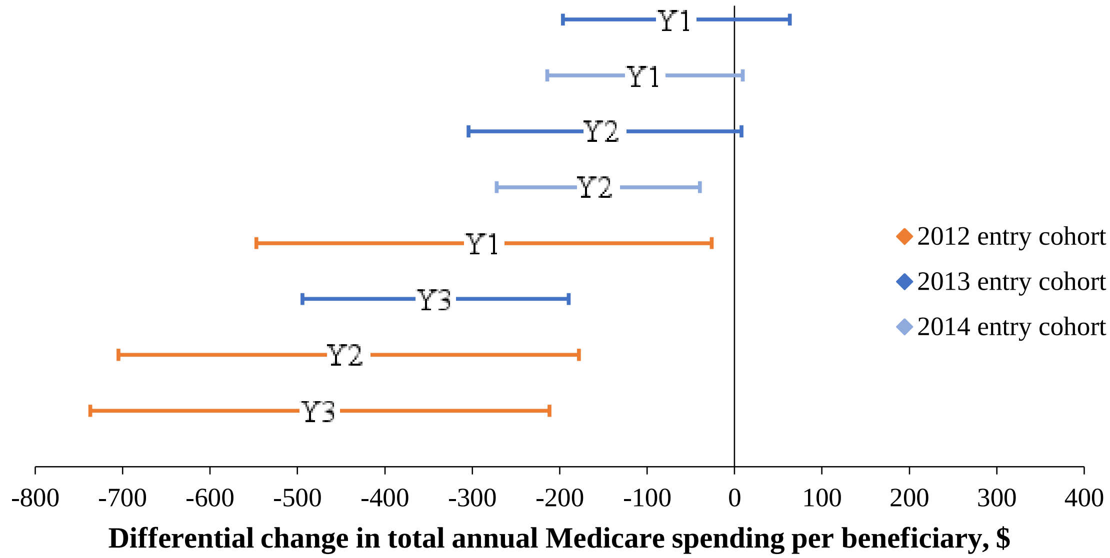
Figure S4. Results of falsification tests hypothetically treating pre-entry years as post-entry years for independent physician group ACOs



Differential changes in total annual Medicare spending per beneficiary over the pre-entry period for entry cohorts of physician group ACOs (from hypothetical pre-entry period [2009 or 2009-2010] to hypothetical first, second, or third post-entry year), $

Figure S5. Differential changes in total Medicare spending for physician group ACO patients from pre-entry period to each post-entry year, by entry cohort and number of years of participation

A. Independent physician group ACOs



REFERENCES

1. McWilliams JM, Hatfield LA, Chernew ME, Landon BE, Schwartz AL. Early Performance of Accountable Care Organizations in Medicare. N Engl J Med 2016;374:2357-66.

2. Li F, Morgan KL, Zaslavsky AM. Li F, Morgan KL, Zaslavsky AM. Balancing covariates via propensity score weighting. 2016.

3. McWilliams JM. Changes in Medicare Shared Savings Program Savings from 2013 to 2014. JAMA 2016;316:1711-13.